Northern glacial refugia for the pygmy shrew (*Sorex minutus*) in Europe revealed by phylogeographic analyses and species distribution modelling.

Rodrigo Vega, Camilla Fløjgaard, Andrés Lira-Noriega, Yoshinori Nakazawa, Jens-Christian Svenning and Jeremy B. Searle.

R. Vega (rrv9@cornell.edu), Department of Biology, University of York, PO Box 373, York YO10 5YW, United Kingdom. (Present address of R. V.: Department of Entomology, Comstock Hall 5123, Cornell University, Ithaca, NY 14853, USA). – C. Fløjgaard, Ecoinformatics and Biodiversity Group, Department of Biological Sciences, Aarhus University, Ny Munkegade 114, Aarhus C and Department of Wildlife Ecology and Biodiversity, National Environmental Research Institute, Aarhus University, Grenaavej 14, DK-8410 Rønde, Denmark. – A. Lira-Noriega and Y. Nakazawa, Natural History Museum and Biodiversity Research Center, The University of Kansas, Lawrence, Kansas 66045, USA.– J.-C. Svenning, Ecoinformatics and Biodiversity Group, Department of Biological Sciences, Aarhus University, Ny Munkegade 114, Aarhus C., Denmark – J.B. Searle, Department of Biology, University of York, PO Box 373, York YO10 5YW, United Kingdom.

Abstract

The Southern European peninsulas (Iberian, Italian and Balkan) are traditionally recognized as glacial refugia from where many species colonized Central and Northern Europe after the Last Glacial Maximum (LGM). However, evidence that some species had more northerly refugia is accumulating from phylogeographic, palaeontological and palynological studies, and more recently from species distribution modelling (SDM), but further studies are needed to test the idea of northern refugia in Europe. Here, we take a rarely implemented multidisciplinary approach to assess if the pygmy shrew (*Sorex minutus*), a widespread Eurasian mammal species, had northern refugia during the LGM, and if these influenced its postglacial geographic distribution. First, we evaluated the phylogeographic and population expansion patterns using mtDNA sequence data from 123 pygmy shrews. Then, we used SDM to predict present and past (LGM) potential distributions using two different training data sets, two different algorithms (Maxent and GARP) and climate reconstructions for the LGM with two different general circulation models. An LGM distribution in the southern peninsulas was predicted by the SDM approaches, in line with the occurrence of lineages of *S. minutus* in these areas. The phylogeographic analyses also indicated a widespread and strictly Northern-Central European lineage, not derived from southern peninsulas, and with a postglacial population expansion signature. This was consistent with the SDM predictions of suitable LGM conditions for *S. minutus* occurring across Central and Eastern Europe, from unglaciated parts of the British Isles to much of the Eastern European Plain. Hence, *S. minutus* likely persisted in parts of Central and Eastern Europe during the LGM, from where it colonized other northern areas during the late-glacial and postglacial periods. Our results provide new insights into the glacial and postglacial colonization history of the European mammal fauna, notably supporting glacial refugia further north than traditionally recognized.

41 **Introduction**

42 During the Quaternary ice ages substantial areas of Northern Europe were covered by ice sheets
43 while permafrost existed in large areas of Central Europe, which restricted the distribution of many
44 temperate and warm-adapted species to the three Southern European peninsulas of Iberia, Italy
45 and the Balkans at the Last Glacial Maximum (LGM; Hewitt 2000). These species are interpreted
46 to have recolonized Central and Northern Europe from these traditionally recognized southern
47 glacial refugia in response to the late-glacial and postglacial warming (Taberlet et al. 1998, Hewitt
48 2000). Therefore, southern glacial refugia and the northward postglacial recolonization of Central
49 and Northern Europe from these areas has become an established biogeographical paradigm
50 (Hewitt 2000).

51 Other studies have, however, provided palaeontological, palynological and phylogeographic
52 evidence that glacial refugia for some temperate and boreal species existed further north than the
53 traditionally recognized Southern European refugia, implying a more complex pattern of glacial
54 survival and postglacial recolonization: Fossils of temperate mammal species dated to the LGM
55 (albeit rarely small mammals) have been described for a number of sites in Central Europe,
56 sometimes in co-occurrence with cold-adapted Pleistocene faunal elements (Sommer and
57 Nadachowski 2006). Macrofossil charcoal (organic plant material ≥ 2 mm in diameter) of coniferous
58 and broad-leaved trees dating to the Upper Palaeolithic has been found in several sites in Austria
59 (42 – 23 Kya), Czech Republic (29 – 24.5 Kya), Croatia (27.8 – 10.8 Kya) and Hungary (31.5 –
60 16.5 Kya), suggesting that these regions were also refugial areas for temperate deciduous species
61 (Willis and van Andel 2004, Magri et al. 2006). Palynological records have shown European beech
62 (*Fagus sylvatica*) pollen in several sites in Central Europe between the late glacial and postglacial
63 (15 – 10 Kya), and have shown that none of the three traditional refugial areas was the source for
64 Northern-Central European beech populations (Magri et al. 2006). Phylogeographic studies on
65 several small mammals have shown little similarity between Mediterranean and northern
66 populations, and have described genetic clades linking together haplotypes sampled throughout
67 Northern-Central Europe (Bilton et al. 1998, Kotlík et al. 2006). Furthermore, species distribution
68 modelling (SDM) has shown that suitable climatic conditions existed for temperate and boreal
69 species in northern latitudes supporting more northerly refugial areas in Europe (Svenning et al.

2008, Fløjgaard et al. 2009). However, a more comprehensive understanding of the relative importance of southern versus northern refugia in terms of LGM species' ranges as well as for postglacial recolonization is needed.

Here, we use the pygmy shrew, *Sorex minutus* Linnaeus 1766 (Mammalia, Soricomorpha), as a model for studying the persistence of populations in Northern European refugia during the LGM. *Sorex minutus* is widely distributed in the Palaearctic, throughout Europe to Lake Baikal (Siberia), including the three Southern European peninsulas (Hutterer et al. 2008). The species occurs at low density in a wide range of terrestrial habitats with adequate ground cover (Churchfield and Searle 2008). In Southern Europe the distribution becomes patchy and limited to higher altitudes where it occurs with some degree of geographical isolation and differentiation, while in central and northern parts of Europe and in Siberia it is more abundant and populations are more connected and widespread.

Previous phylogeographic studies on *S. minutus* revealed a very widespread and genetically homogeneous 'Northern-Central European and Siberian' lineage, extending from Britain through Central and Northern Europe to Siberia (c. 7000 km), but genetically distinct from the southern lineages in Iberia, Italy and the Balkans (Bilton et al. 1998, Mascheretti et al. 2003, McDevitt et al. in press). These studies suggested that the Northern-Central European lineage persisted and expanded from one or more Central or Eastern European refugia located further north than the traditionally recognized Southern European refugia. However, the size and locations of the possible northern refugia for *S. minutus* could not be assessed precisely.

Species distribution models combine information about species occurrences with environmental (usually climatic) data found across the study region to estimate the present-day geographical distribution of suitable environmental conditions for the species (Guisan and Zimmermann 2000). Then, the set of environmental conditions can be projected to past conditions to identify areas where there were suitable environmental conditions for the species (hindcasting) (Nogués-Bravo 2009), in this case at the LGM. Such SDM-based hindcasting has not been integrated into the previous phylogeographic studies on *S. minutus*, and the genetic data for central and eastern regions of Europe and in Siberia have been rather incomplete. This makes it difficult to determine the importance of these regions for the LGM distribution of the species, its

99 postglacial colonization history and its present-day genetic structure. Moreover, the inference of
100 glacial refugia based solely on phylogeographic analyses can be obscured by the extinction of
101 genetic variants, incomplete sampling and large-scale range shifts of the species (Waltari et al.
102 2007). Hence at this point, although the previous phylogeographic studies suggested the existence
103 of northern glacial refugia for *S. minutus*, the size and geographic spread of these refugia as well
104 as their role in the postglacial range dynamics of the species remain unclear.

105 The purpose of this study is to assess the distribution of *S. minutus* during the LGM based
106 on a multidisciplinary approach using more detailed mtDNA-based phylogeographic analyses than
107 conducted hitherto and including SDM-based hindcasting. Only a few studies have tried to
108 estimate potential northern refugial areas in this way, despite the stronger inference allowed by
109 these independent and highly complementary approaches (Waltari et al. 2007).

110 We assessed the following specific study questions: Would a more detailed
111 phylogeographic analysis also detect a distinctive 'Northern-Central European and Siberian'
112 lineage as has been previously found? Would this widespread lineage present a genetic signature
113 of population expansion? Would different SDM-based hindcasting approaches predict suitable
114 LGM conditions for *S. minutus* not only in the Southern European peninsulas, but also further
115 north, consistent with northern refugia? Would the combined phylogeographic and SDM approach
116 allow us to estimate more precisely the geographic locations of northern refugia for *S. minutus*, as
117 well as determine their potential role for its postglacial range dynamics? From the population
118 expansion characteristics, how did the refugial populations colonize their current ranges? Finally,
119 are the rather scant fossil data for *S. minutus* consistent with our phylogeographic and
120 distributional findings?

121 This study sheds light on the spatial variation of the genetic diversity within the widespread
122 distribution of *S. minutus*, its postglacial population expansion and colonization of Europe from
123 northern refugia, and contributes towards an emerging new synthesis of the full-glacial distributions
124 of the European biota. The nature of northern refugia also has important implications for the
125 understanding of their biogeographic roles as sources of genetic diversity, areas of speciation,
126 identification of conservation units and preservation of species, particularly in response to future
127 climate change (Kotlík et al. 2006, Provan and Bennett 2008).

Materials and Methods

Phylogeographic analyses

Samples and laboratory procedures

In total, 123 individuals of *S. minutus* from Europe and Siberia were used for the phylogeographic analysis of the mitochondrial cytochrome *b* (cyt *b*) gene. Sixty-six *S. minutus* cyt *b* sequences were obtained from Genbank (AB175132: Ohdachi et al. 2006; AJ535393-AJ535457: Mascheretti et al. 2003). Fifty seven out of the 123 samples of *S. minutus* were obtained from Northern-Central Europe during fieldwork and from museum collections (see Acknowledgements) to increase the molecular data and to provide a more detailed analysis of this region. A sequence of *S. volnuchini* was used as outgroup (AJ535458: Mascheretti et al. 2003).

Genomic DNA was extracted using a commercial kit (Qiagen). Partial cyt *b* sequences were obtained by PCR using two primer pairs that amplified ca. 700 bp of overlapping fragments. PCR amplification was performed in a 50 µl final volume: 1X Buffer, 1 µM each primer, 1 µM dNTP's, 3 mM $MgCl_2$ and 0.5 U Platinum *Taq* Polymerase (Invitrogen), with cycling conditions: 94 ℃ for 4 min, 40 cycles at 94 ℃ for 30 s, 55 ℃ for 30 s and 72 ℃ for 45 s, and a final elongation step at 72 ℃ for 7 min. Purification of PCR products was done with a commercial kit (Qiagen) and sequenced (Macrogen and Cornell University Core Laboratories Center).

Sequence and phylogenetic analyses

Cyt *b* sequences were edited in BioEdit 7.0.9.0 (Hall 1999) and aligned by eye. For the construction of phylogenetic trees, the model of evolution that best fitted the molecular data was searched using MrModeltest 2.3 (Nylander 2004) using the minimum Akaike Information Criteria value. The substitution model supported was the General Time Reversible with specified substitution types (*A*–*C* = 0.3663, *A*–*G* = 17.4110, *A*–*T* = 1.0216, *C*–*G* = 2.1621, *C*–*T* = 13.0604, *G*–*T* = 1.0), proportion of invariable sites (0.5332), gamma shape parameter (0.9799) and nucleotide frequencies (*A* = 0.2750, *C* = 0.2996, *G* = 0.1382, *T* = 0.2872).

156 The phylogenetic relationships within *S. minutus* were inferred by Neighbour-Joining (NJ),
157 Maximum Likelihood (ML) and Bayesian analysis using PAUP* 4.0b10 (Swofford 2000), PhyML 3.0
158 (Guindon and Gascuel 2003) and MrBayes 3.1 (Huelsenbeck and Ronquist 2001), respectively.
159 Confidence for the phylogenetic relationships in NJ and ML was assessed by bootstrap replicates
160 (10,000 and 500 replicates, respectively). For the Bayesian analysis, two independent runs were
161 performed with 10 million generations and 5 chains each, a sampling frequency every 1000
162 generations, a temperature of 0.1 for the heated chain and checking for convergence. Trees were
163 summarized after a burn-in value of 2500 to obtain the posterior probabilities of each phylogenetic
164 branch.
165 Phylogenetic networks provide an explicit graphic representation of evolutionary history
166 between sequences in which taxa are represented as nodes and their evolutionary relationships
167 are represented by edges. Most internal nodes represent ancestral states from which more recent
168 and peripheral nodes derive (Avise 2000). A parsimony phylogenetic network of cyt *b* haplotypes
169 was constructed using the software Network 4.5.1.0 (Fluxus-Engineering) with a median-joining
170 algorithm and a greedy FHP genetic distance calculation method. The median joining algorithm
171 identifies groups of haplotypes and introduces hypothetical (non-observed) haplotypes to construct
172 the parsimony network.
173
174 *Genetic and statistical analyses*
175 Standard sequence polymorphism indices (number of haplotypes, polymorphic sites and
176 parsimony informative sites) and genetic diversity values (π, nucleotide diversity ±SD; *h*, haplotype
177 diversity) were estimated using Arlequin 3.11 (Excoffier et al. 2005).
178 Population expansion was examined for both the full dataset (Eurasia) and for the
179 'Northern-Central European and Siberian' lineage using DnaSP 5.00.04 (Librado and Rozas 2009).
180 In each case a mismatch distribution (distribution of the number of differences between pairs of
181 haplotypes) was estimated to compare the demography of the populations with the expectations of
182 a sudden population expansion model (Rogers and Harpending 1992). The raggedness index (*rg*),
183 which measures the smoothness of the observed distribution, was computed and the statistical

184 validity of the estimated expansion model was tested by a parametric bootstrap approach as a sum
185 of square deviations (SSD) between the observed and the expected mismatch (Schneider and
186 Excoffier 1999) using Arlequin (10,000 replicates). Three other tests for population expansion were
187 performed in DnaSP using coalescent simulations to test for statistical significance (10,000
188 replicates): R^2 test of neutrality, based on the difference of the number of singleton mutations and
189 the average number of nucleotide differences (Ramos-Onsins and Rozas 2002); Fu's *Fs*, a statistic
190 based on the infinite-site model without recombination that shows large negative *Fs* values when
191 there has been a demographic population expansion (Fu 1997); Tajima's *D*, a test for selective
192 neutrality based on the infinite-site model without recombination where significant values appear
193 from selective effects but also from factors such as population expansion, bottleneck or
194 heterogeneous mutation rates (Tajima 1989).
195
196 **Species distribution modelling**
197 Important discrepancies in the prediction of the potential distribution of a particular species arise
198 from differences in data sample size (Stockwell and Peterson 2002, Wisz et al. 2008),
199 environmental and/or climatic data (Peterson and Nakazawa 2008), and algorithms (Peterson et al.
200 2007, but see Phillips 2008). Also, if the occurrence records used to model the distribution do not
201 adequately sample the environmental requirements of the species, the prediction will not truly
202 reflect its potential geographic distribution (Pearson et al. 2007). Therefore, to ensure the
203 robustness of our findings, we modelled the potential distribution of *S. minutus* in the present and
204 at the LGM using two independent training data sets, two algorithms, namely the maximum entropy
205 algorithm (Maxent; Phillips et al. 2006) and the Genetic Algorithm for Rule-set Prediction (GARP;
206 Stockwell and Noble 1992, Stockwell 1999), and using climate reconstructions for the LGM based
207 on two general circulation models (GCMs). All GIS operations were performed using ArcGIS 9.3
208 (ESRI, Redlands, CA, USA).
209
210 *Species occurrence data*
211 For the first data set, hereafter termed 'data set 1', we used the species records from fieldwork,
212 from two online sources (Global Biodiversity Information Facility, GBIF, and Mammal Networked

213 Information System, MaNIS) and from museum specimens obtained for our study (see
214 Acknowledgements). Most of the data were derived from the following sources: the Atlas of
215 Mammals in Britain (Arnold 1993), the European Environment Agency, the UK National
216 Biodiversity Network, the Highland Biological Recording Group – HBRG Mammals data set and the
217 Ministerio de Medio Ambiente y Medio Rural y Marino (Spain). Low precision occurrences, such as
218 presence data taken from the centroids of atlas grids and falsely georeferenced occurrences (i.e.,
219 offshore and out-of-range locations), were eliminated from this data set. In total, we collected 536
220 high-precision unique latitude-longitude localities, but this data set was geographically biased
221 towards Western Europe and Britain due to differences in sampling effort across the species'
222 distribution range (i.e., there are few species records from Siberia and Southern Europe). In order
223 to correct for sampling bias, we created 25 random subsets from the original data set to limit the
224 number of unique occurrences to ≤ 5 in squares of 5 x 5 degrees distributed across the extent of
225 the geographical analysis (Wisz et al. 2008). This procedure yielded a total of 146 unique localities
226 for each subset which were more evenly distributed.

227 For the second data set, hereafter termed 'data set 2', we used the records from the Atlas
228 of European Mammals (AEM; Mitchell-Jones et al. 1999) which present less geographic bias within
229 Europe, but had a much coarser resolution than data set 1. The AEM uses an approximate equal
230 area grid of 50 x 50 km based on the Universal Transverse Mercator (UTM) projection and the
231 Military Grid Reference System (MGRS). Records of 'species presence' as well as 'presence
232 assumed' (i.e., presence was observed before 1970 and no evidence of later extinction) were
233 included in the study and a total of 1178 data points were used.

234 To ensure transferability of our models, we used a geographically independent test data
235 set. We digitized the Eurasian range map for *S. minutus* (Hutterer et al. 2008) and recorded the
236 species as present in all 50 x 50 km MGRS grid cells within the outline of the range map. Then, we
237 used the part of the range located east of the European study area (for simplicity referred to
238 hereafter as Siberia) only as a test data set (n = 3122 data points). This allowed us to evaluate the
239 performance of the models with both data sets and assess which climatic variables provided the
240 strongest predictive ability in a geographically independent region with relatively LGM-like
241 conditions (Fløjgaard et al. 2009). We used the digitised range map data only for testing, given its

much coarser resolution and uncertain quality compared to the occurrence data from data sets 1 and 2.

Climate data

For the present-day SDM we initially considered the 19 bioclimatic variables from the WorldClim dataset at a spatial resolution of 2.5 minutes <http://www.worldclim.org/>. These climate layers are based on spatially interpolated values of temperature and precipitation gathered from weather stations around the world from 1950-2000 (Hijmans et al. 2005). For the LGM *(*21 kya) we used the climate reconstructions of the same 19 bioclimatic variables based on the CCSM3 (Collins et al. 2006) and MIROC3.2 (Hasumi and Emori 2004) GCMs <http://biogeo.berkeley.edu> at a spatial resolution of 2.5 minutes.

We used the Jackknife procedure implemented in Maxent with the 19 bioclimatic variables on the two data sets to find the best set of predictor variables. We assessed the performance of the models based on the Area Under the Curve (AUC) values of the Receiver Operating Characteristic (ROC) in the independent test region of Siberia. The worst predictor of the whole set of variables was eliminated, a new model was produced using the remaining variables and the process was repeated until all variables were exhausted. We chose the final set of predictors based on parsimony (i.e., with the fewest number of climatic variables) and with the highest AUC value in the independent test region of Siberia.

The final set of predictors comprised the variables Annual Mean Temperature (AMT) and Precipitation of the Warmest Quarter (PWQ); thus, AMT and PWQ were used for estimating the present and LGM distribution of *S. minutus*. These two variables were not highly correlated (r = -0.3550) and models that included only these yielded higher or almost equal AUC values than models that included only one or more variables in combination with AMT and PWQ. In addition, these variables are biologically meaningful for *S. minutus* considering its broad distribution in Northern-Central Europe and Siberia and habitat preference for damp and temperate areas (Churchfield and Searle 2008, Hutterer et al. 2008). The modelling was performed with data sets 1 and 2 as inputs in Maxent and GARP, and all models were evaluated on the geographically independent (extrinsic) test data from Siberia. For data set 1 we made models with all 25 subsets.

Finally, all models were projected onto the two LGM climate reconstructions to identify the potential distribution of *S. minutus*.

Modelling algorithms

To assess the variation in the outcome of model predictions due to differences in modelling algorithms, we chose to use Maxent and GARP. Maxent has been shown to perform very well in comparative studies of species distribution modelling compared to GARP (Elith et al. 2006, Phillips and Dudík 2008, Elith and Graham 2009, but see also Peterson et al. 2008), while GARP has been shown to perform better than Maxent in transferability studies (Peterson et al. 2007, but see also Phillips 2008). Ultimately, the performance of each algorithm may be properly compared using the corresponding thresholding during model evaluation, since their predictions are not given in the same scale (Peterson et al. 2008).

To evaluate the accuracy of our models, the empirical AUC values were compared against the AUC values of 1000 random models, as implemented in Peterson et al. (2008), using the data from the test region. AUC ROC values are expressed as the ratio of the area under the observed curve (i.e., the overall area for which each algorithm predicts as present) to the area under the line that defines a random expectation; consequently, the AUC values are expected to be larger than one as the model departs from the random expectation (Peterson et al. 2008).

Maxent is a machine-learning technique based on the principle of maximum entropy that fits a probability distribution to the environmental conditions at the locations where a species has been observed (Phillips et al. 2004, Phillips et al. 2006). When implemented with ecologically meaningful sets of predictor variables, Maxent produced similar estimates for the locations of glacial refugia as Bioclim, another commonly used, but simpler, modelling technique (Svenning et al. 2008, Fløjgaard et al. 2009). We used the default settings in Maxent 3.2.1 <http://www.cs.princeton.edu/~schapire/maxent/> with background data limited to Eurasia as described in the species occurrence data section. We converted the continuous logistic output from Maxent into a binary map of predicted suitable environmental conditions for *S. minutus* using the

298 maximum test sensitivity and specificity threshold because it optimized the correct discrimination of
299 presences and pseudoabsences in the test data.
300 GARP is a genetic algorithm that produces a set of rules that describe the non-random
301 association between environmental variables and occurrence data (Stockwell and Noble 1992,
302 Stockwell 1999). First, the algorithm creates a set of rules based on four basic types (bioclimatic,
303 atomic, negated and logistic regression rules), their individual predictive accuracy is calculated and
304 only those rules with the highest predictive accuracy are retained in the model. The overall
305 performance of the model is evaluated using a subset of presence points. Then, a second
306 generation of rules is produced via the random modification of the previous generation rules, their
307 predictive accuracy is calculated and only those with the highest accuracy are included in the
308 model. Finally, the overall performance of the model is re-evaluated and the process of creation,
309 evaluation and inclusion of rules is repeated until a maximum number of iterations is reached
310 (1000 in this case), or until performance values no longer change appreciably from one iteration to
311 the next (convergence parameter of 1%). We used the version of DesktopGarp as implemented in
312 openModeller version 1.0.9 <http://openmodeller.sf.net> using the default parameters (Anderson et
313 al. 2003). We converted the continuous output into a binary map of predicted occurrence of the
314 suitable conditions for *S. minutus* by assigning a value of 1 for the model values that corresponded
315 to 10% or more of the testing points.
316
317 **Results**
318
319 **Phylogeographic analysis**
320 *Sequence analysis and phylogenetic reconstructions*
321 A partial sequence of 1110 bp from the *S. minutus* cyt *b* was analysed. One hundred and twelve
322 haplotypes were obtained, from which 46 were newly described and deposited in GenBank
323 (accession numbers: GQ494305-GQ494350). There were 894 invariable and 216 variable
324 positions, from which 137 were parsimony informative.
325 All the phylogenetic analyses revealed five distinct lineages (Fig. 1). Samples from the
326 Mediterranean peninsulas clustered in three lineages, namely the ‘Iberian’, ‘Italian’ and ‘Balkan’

groups, corresponding to their geographical origin. There was also a well supported 'Pyrenean' lineage with samples from Andorra and Ireland. Samples from Northern-Central Europe and Siberia clustered together forming a geographically widespread lineage that did not include any individuals from the southern peninsulas, hereafter named as the 'Northern-Central European' lineage. This lineage was composed of 105 sequences (94 haplotypes) with 940 invariable and 170 variable positions, from which 92 were parsimony informative.

The phylogenetic network of the Northern-Central European lineage presented a star-like pattern with three most central haplotypes, named A, B and C, separated by only one mutational step from each other and from which all other sequences derived (Fig. 2). The other phylogroups from the southern peninsulas were much more distantly related and separated by several mutations (data not shown). The central haplotypes A and B were entirely composed of samples from the Netherlands (three and two individuals, respectively), while the third central haplotype (C) belonged to a Central Ukrainian specimen from the locality Tishki (50° 6.27' N, 33° 6.39' E). There was an apparent geographical subdivision of the samples that were connected to these three central haplotypes (Fig. 2). Only haplotypes from Great Britain and the Netherlands were directly connected to A. Several haplotypes from different countries of Northern and Central Europe were connected to B, also including some haplotypes from Great Britain and the Netherlands, but there were no haplotypes from Eastern Europe or Siberia (except for one sample from Ukraine ambiguously connected to B and C). Haplotypes from Northern, Central and Eastern Europe and Siberia were all directly connected to C, but there were no samples from countries further west than Germany. However, the support for these subdivisions was not strong: equally parsimonious explanations (loops) appeared in the central part of the network between B and C, and there was no supported sub-structure within the Northern-Central European lineage in the phylogenetic trees.

Genetic and statistical analyses

The whole Eurasian sample presented a nucleotide diversity $\pi = 0.0109 \pm 0.0055$, and a haplotype diversity $h = 0.9983$. The Northern-Central European lineage had a nucleotide diversity $\pi = 0.0067$

354 ± 0.0035, and a haplotype diversity h = 0.9980. Genetic diversity values were not calculated for the
355 Southern European lineages because of small sample size.

356 The mismatch distribution of the whole dataset (Eurasia) was bi-modal, consistent with
357 pairwise differences between sequences belonging to the same and different lineages (Fig. 3a).
358 The mismatch distribution of the Northern-Central European lineage showed a unimodal
359 distribution that, visually, fitted almost perfectly over the expected values for a population
360 expansion model (Fig. 3b). There was an observed mean of 7.382 pairwise differences with a
361 variance of 8.152. The goodness of fit test showed no significant differences between the observed
362 and expected values under a sudden expansion model for the Northern-Central European lineage
363 (SSD = 0.0004, $p_{SSD} > 0.05$; rg = 0.0082, $p > 0.05$). Negative and significant Tajima's D (D = -
364 2.5721, $p < 0.001$) and Fu's Fs (Fs = -24.8437, $p < 0.001$) showed departures from neutrality also
365 consistent with a sudden population expansion. Moreover, the R^2 test of neutrality also showed
366 that the Northern-Central European lineage gave a genetic signature consistent with a sudden
367 population expansion (R^2 = 0.0180, $p < 0.001$). The rest of the sequences and lineages that
368 belonged to the more distantly related Southern European lineages (Iberian, Italian and Balkan
369 peninsulas) and the Pyrenees were not analysed because of small sample size.
370
371 **Species distribution modelling**
372 *Predicted present distribution*
373 Species distribution models from Maxent matched the reported distribution of the species (Fig. 4a,
374 c). The models also predicted suitable climatic conditions outside the reported distribution of the
375 species especially in two regions, the Asia Minor-Caucasus region and in the Far East (Fig. 4a, c).
376 The predicted present distribution of *S. minutus* with GARP was very similar to that of Maxent, it
377 also matched the reported distribution and the predicted suitable climatic conditions in the Asia
378 Minor-Caucasus region and in the Far East (Fig. 4b, d).

379 All Maxent and GARP models were accurate in the test region, with AUC values for both
380 data sets higher than null expectations ($p < 0.001$; mean AUC_{MAXENT} = 1.24 ±0.021 and mean
381 AUC_{GARP} = 1.049 ±0.007 for data set 1, and mean AUC_{MAXENT} = 1.249 ±0.011 and mean AUC_{GARP} =
382 1.032 ±0.005 for data set 2).

383

384 *Predicted LGM distribution*

385 With the two data sets and GCMs, Maxent and GARP predicted suitable LGM climatic conditions in
386 the Southern European peninsulas (Fig. 4e – l), concordant with southern refugia. In general,
387 suitable LGM conditions with the two data sets, GCMs and algorithms were also predicted north of
388 the southern refugia, particularly throughout Central Europe, most of Eastern Europe, Southern
389 Poland, Eastern and Southern Ukraine, the Crimea peninsula and the Caucasus. With Maxent, the
390 LGM predictions differed little between data sets or between GCMs, and there were predicted
391 suitable conditions in Central and Eastern Europe close to the ice sheet (Fig. 4e, g, i, k). With
392 GARP, predictions differed between GCMs: more restricted suitable conditions in Central and
393 Eastern Europe were predicted with CCSM3 (Fig. 4f, h) than with MIROC3.2 (Fig. 4j, l), but
394 predictions did not differ much between data sets. The most restricted predictions (using GARP
395 with CCSM3) still showed suitable climatic conditions in Southern Ireland, Central and Southern
396 France, western parts of Switzerland, a few regions north of the Balkans, the Crimea peninsula
397 and the Caucasus.

398

399 **Discussion**

400

401 **Northern glacial refugia revealed by a combined approach**

402 *Sorex minutus* is considered a temperate species, but it is also latitudinally distributed above 60°
403 North (i.e., near the Arctic Circle) and altitudinally above 2000 m in regions with permafrost and
404 harsh winters (Mitchell-Jones et al. 1999, Hutterer et al. 2008). Northern non-arctic species like *S.*
405 *minutus* could have persisted in high latitude refugia in Europe during the LGM, north of the
406 traditionally recognized Mediterranean refugial areas (Stewart and Lister 2001). This could have
407 been a result of their ecological traits (notably cold tolerance) and biogeographical characteristics
408 that may have determined their response to the glaciations (Bhagwat and Willis 2008). *Sorex*
409 *minutus* is, therefore, a suitable model organism for exploring the controversial hypothesis of
410 'northern' glacial refugia.

411 The general concordance of the phylogeographic analyses with the predicted LGM
412 distributions based on species distribution modelling and the concordance between models
413 suggest that we have obtained robust results concerning the LGM distribution of *S. minutus*. Our
414 phylogeographic analyses provided evidence for a distinct lineage in Northern-Central Europe, with
415 additional lineages in the Iberian, Italian and Balkan peninsulas in Southern Europe. First, the
416 absence of southern haplotypes in Northern-Central Europe supports the hypothesis that the
417 southern peninsulas were areas of endemism and differentiation for *S. minutus*, but not for
418 northward colonization (Bilton et al. 1998), i.e., the current populations in Northern-Central Europe
419 were not derived from LGM populations in the traditional Southern European refugia. Second, the
420 Northern-Central European lineage showed a strong signature of population expansion supported
421 by the mismatch distribution and population expansion tests. Finally, ancestral haplotypes in a
422 phylogenetic network can be identified by their central or internal position from where the
423 peripheral, more recent, haplotypes are derived, by the number of haplotypes that arise from them
424 and by their abundance (Avise 2000). The phylogenetic network of the Northern-Central European
425 lineage showed a star-like pattern with three ancestral haplotypes from distant regions in Central
426 and Eastern Europe (the Netherlands and Ukraine). This pattern was also consistent with a
427 widespread LGM distribution and congruent with the hypothesis of persistence and postglacial
428 expansion from northern glacial refugia.
429 The phylogeographic pattern that we observe here did not arise from the low sample size in
430 Southern Europe: the few samples from southern peninsulas belonged to lineages differentiated by
431 a large number of mutation steps from the Northern-Central European lineage; if Northern-Central
432 Europe had been colonized from Southern Europe we would have found Northern-Central
433 European samples clustering within southern lineages, not forming a separate lineage. Moreover,
434 a phylogeographic study on *S. minutus* using the mitochondrial Control Region and Y-chromosome
435 introns with more samples from southern peninsulas showed a similar pattern (McDevitt et al. in
436 press). Nevertheless, further sampling in southern regions and the use of other molecular markers
437 is desirable to investigate the genetic variation and population expansion events within
438 Mediterranean peninsulas, and for the determination of contact zones among lineages.

439 We did not use the mismatch distributions to date the population expansion for the
440 Northern-Central European lineage because of the lack of a suitable mutation rate for cyt *b* in *S.*
441 *minutus.* Previous studies on *Sorex* have used mismatch distributions for molecular dating (e.g.
442 Ratkiewicz et al. 2002), but with mutation rates that may not be suitable over short time frames (Ho
443 et al. 2005).
444 The modelling approaches predicted successfully the wide present-day distribution of *S.*
445 *minutus* in Eurasia. Therefore, we consider our SDM approaches as giving realistic estimates of
446 the area with suitable climatic conditions for our species and of its potential LGM distribution. A
447 third model using Bioclim with SDM data sets 1 and 2 also resulted in very similar present-day and
448 LGM distributions for *S. minutus* (data not shown). The potential LGM distributions predicted by our
449 SDM approaches not only included the traditionally recognized southern refugia, but also a wide
450 area across Central and Eastern Europe, from the unglaciated parts of Southern Ireland and
451 Britain to most of the Central and Southeast European (or Russian) Plain. In particular, the
452 predicted LGM distribution throughout Central and Eastern Europe encompasses suggested
453 northern refugial areas based on palaeontological and palynological data for other temperate and
454 boreal species (Willis et al. 2000, Willis and van Andel 2004, Magri et al. 2006, Sommer and
455 Nadachowski 2006). Thus, the Northern-Central European lineage could have persisted in various
456 parts of this wide area during the LGM according to the phylogeographic and the SDM
457 approaches.
458 We note that the Central and Eastern European LGM distribution was similar with both data
459 sets, particularly when using Maxent (with both GCMs) and when using GARP with MIROC3.2,
460 even though we used very different species records. However, the LGM distributions when using
461 GARP were more widespread to the north with MIROC3.2 than with CCSM3 GCMs, which could
462 represent variations due to modelling algorithms and GCMs. Also, the predicted present-day
463 suitable climatic conditions outside the reported distribution of *S. minutus* in the Asia Minor-
464 Caucasus region and in the Far East probably reflect competitive or speciation processes rather
465 than an inaccurate estimation of the suitable climatic conditions. In Asia Minor-Caucasus, *S.*
466 *minutus* is replaced by the closely related sister species *S. volnuchini*, while in the Far East many
467 other *Sorex* species occur including similar-sized species such as *S. gracillimus.*

468 The predicted LGM distribution of *S. minutus* appears to be continuous throughout Europe;
469 however, lineage diversification is still plausible: First, the present distribution of *S. minutus* also
470 appears to be continuous but it is affected by landscape features, not evident at the geographic
471 resolution given, which could have subdivided the species range. Therefore, it could be expected
472 that landscape features at the LGM also affected the distribution of *S. minutus*. Second, the
473 estimation of the extent of ice sheets in mountainous areas is not precise, so it may be expected
474 that the Iberian and Italian populations remained isolated from the rest of Europe by ice sheets
475 covering the Pyrenees and the Alps, respectively, while the heterogeneous landscape in the
476 Balkans could have been responsible for the limited distribution of the genetic lineage there. Also,
477 different genetic variants could have arisen within regions and could have been maintained there
478 selectively reducing further spread into contiguous regions. Another explanation could be that
479 interspecific competition and/or other non-climatic conditions subdivided the potentially continuous
480 LGM distribution.
481
482 **Insights into postglacial colonization**
483 The predicted distribution for *S. minutus* in the Iberian, Italian and Balkan peninsulas presumably
484 corresponds to the refugial areas where the southern genetic lineages persisted during the LGM.
485 The Pyrenean lineage, here represented by a limited number of Andorran and Irish samples, could
486 have persisted during the LGM in Central and South-western France and even in unglaciated
487 areas in Southern Ireland, as shown by our SDM models. However, genetic studies support a more
488 recent origin of the Irish pygmy shrew, transported there by humans during the Holocene
489 (Mascheretti et al. 2003, McDevitt et al. 2009, A.D. McDevitt, V.R. Rambau, R. Vega and J.B.
490 Searle, unpublished data). Further molecular sampling in Southern Europe is desirable to
491 determine the extent of the geographic distribution of the lineages found there and the contact
492 zones between them.
493 Considering the phylogenetic network for the Northern-Central European lineage, the three
494 central (ancestral) haplotypes were located in or near regions where the SDM approaches
495 predicted a potential LGM distribution for *S. minutus*. These results imply that *S. minutus* was not
496 dependent on amelioration of the climate at the end of the last glaciation to colonize Northern-

497 Central Europe from southern refugia; instead, it was already present. As the ice sheets retreated
498 and the climate improved, the range of *S. minutus* expanded from northern refugia colonizing the
499 rest of Northern-Central Europe. For example, Scandinavian and the Baltic regions were most
500 likely colonized by pygmy shrews from Eastern Europe, not from the west or from southern
501 peninsulas. Thus, the phylogenetic network shows that sequences from Norway, Finland and
502 Lithuania group closely with the Ukrainian central haplotype, which according to the SDM
503 modelling could have survived the LGM *in situ* on the East European Plain. Likewise, the genetic
504 similarity of samples from the Netherlands and Britain, in comparison to those elsewhere, suggests
505 that the British pygmy shrew originated from populations in the vicinity of the Netherlands, reaching
506 Britain over the landbridge with continental Europe. An alternative explanation is that *S. minutus*
507 persisted in the unglaciated regions of Southern Britain (as predicted by several of our SDM
508 approaches) which were geographically connected and genetically similar to populations in
509 continental Europe during the LGM. Whatever the explanation, as ice sheets retreated, *S. minutus*
510 belonging to the Northern-Central European lineage was able to colonize the northern parts of
511 mainland Britain.
512
513 **Further support from fossils and phylogeographic analyses**
514 Northern refugia in Central Europe and further east, north of the traditional Mediterranean refugia,
515 have been hypothesized in phylogeographic analyses for a number of small mammal species other
516 than the pygmy shrew, including the field vole (*Microtus agrestis*, Jaarola and Searle 2002), bank
517 vole (*Clethrionomys glareolus*, Deffontaine et al. 2005, Kotlík et al. 2006), root vole (*Microtus*
518 *oeconomus*, Brunhoff et al. 2003), common vole (*Microtus arvalis*, Heckel et al. 2005) and the
519 common shrew (*Sorex araneus*, Bilton et al. 1998, Yannic et al. 2008). For bank voles, root voles,
520 field voles and common voles, predictions of their potential LGM distribution based on SDM were
521 also consistent with northern refugia (Fløjgaard et al. 2009).
522 Most of the phylogeographic studies point to the Carpathians as a likely northern refugial
523 area, but a refugium in this area could have included broader regions of Hungary, Slovakia, Czech
524 Republic, Moldova and Poland, supported by the occurrence of temperate mammal fossil records
525 in the area (Sommer and Nadachowski 2006) and by our results. Also, the region of the Dordogne

526 in South-western France was situated outside the LGM permafrost area and has temperate
527 mammal fossil records dated to the end of the LGM. Therefore, it has been suggested as another
528 likely refugium north of the traditionally recognized southern refugia (Sommer and Nadachowski
529 2006), further supported by our findings.

530 In addition, there are a few but important fossil records of *S. minutus* from several localities
531 north of the southern refugia, radiocarbon dated close to the LGM or earlier (S3P Faunal Database
532 <http://www.esc.cam.ac.uk/research/research-groups/oistage3/stage-three-project-database-
533 downloads>). These fossil remains have been found in sites in France (26 Kya), Belgium (38 – 40
534 Kya), Germany (23 – 29 Kya) and Hungary (20 – 22 Kya).

535 In conclusion, a wide northern LGM distribution for *S. minutus* is supported by the
536 combined use of a phylogeographic and species distribution modelling approach. The SDM
537 methodologies provide evidence for a Central and Eastern European LGM distribution of *S.*
538 *minutus*, where the Northern-Central European lineage could have been distributed. Additionally,
539 the SDM approaches reveal potential LGM distributions for *S. minutus* in southern refugia,
540 consistent with the lineages present in those regions. The phylogeographic analyses, however,
541 indicate that the southern refugia were not the postglacial source of the current and widespread
542 Northern-Central European populations. The other phylogeographic and SDM studies on small
543 mammals, mammal and plant fossil records, and *S. minutus* fossil remains presented here provide
544 additional evidence consistent with or directly supportive of our findings.

545 Our results contribute to the understanding of persistence and colonization from glacial
546 refugia further north than traditionally recognized. They also provide new insights into the location
547 and importance of refugial areas for the persistence of populations and genetic lineages during
548 climate change. The use of *S. minutus* as a model exemplifies how the combined use of
549 phylogeography and species distribution modelling can be applied to understand present-day
550 biodiversity patterns, and can predict and test the past distribution of species to gain insight into
551 the colonization patterns, differentiation and biogeography of species.
552

Acknowledgements – Specimens and species records of *Sorex minutus* were made available by several museums and we acknowledge the help of the curators from the following institutions: Department of Ecology and Evolution (Université de Lausanne, Switzerland), Univerza na Primorskem and Research Centre of Koper (Slovenia), Natuurhistorisch Museum (Rotterdam, the Netherlands), Dipartimento di Ecologia (Università della Calabria, Italy), Museo di Anatomia Comparata and Museo di Zoologia 'La Sapienza' (Università di Roma, Italy) and Natuurmuseum Brabant (Tilburg, the Netherlands). We are very grateful for the tissue samples provided by Boris Kryštufec, Allan McDevitt, Glenn Yannic, Jacques Hausser, Jan Zima, Fríða Jóhannesdóttir, Holger Bruns, Peter Borkenhagen and Petr Kotlík. We thank David Nogués-Bravo and two anonymous referees for their valuable comments. Bayesian analyses were run at the Computational Biology Service Unit from Cornell University which is partially funded by Microsoft Corporation. We gratefully acknowledge financial support to R. Vega (181844) and A. Lira-Noriega (189216) from CONACyT (México), and to J.-C. Svenning from the Danish Natural Science Research Council (grant 272-07-0242).This work represents the fruits of the discussion of our work presented at the 4th Biennial Conference of the International Biogeography Society (8–12 January 2009, Mérida, Yucatán, México).

References

Anderson, R.P. et al. 2003. Evaluating predictive models of species' distributions: criteria for selecting optimal models. – Ecol. Model. 162: 211–232.

Arnold, H.R. 1993. Atlas of mammals of Britain. – Her Majesty's Stationary Office, London.

Avise, J.C. 2000. Phylogeography: the history and formation of species. – Harvard Univ. Press.

Bhagwat, S.A. and Willis, K.J. 2008. Species persistence in northerly glacial refugia of Europe: a matter of chance or biogeographical traits? – J. Biogeogr. 35: 464–482.

Bilton, D.T. et al. 1998. Mediterranean Europe as an area of endemism for small mammals rather than a source for northwards postglacial colonization. – Proc. R. Soc. B. 265: 1219–1226.

Brunhoff, C. et al. 2003. Holarctic phylogeography of the root vole (*Microtus oeconomus*): implications for late Quaternary biogeography of high latitudes. – Mol. Ecol. 12: 957–968.

581 Churchfield, S. and Searle, J.B. 2008 Pygmy shrew *Sorex minutus*. – In: Harris, S. and Yalden,
582 D.W. (eds.), Mammals of the British Isles: handbook. Mammal Society, pp. 267–271.
583 Collins, W.D., et al. 2006. The Community Climate System Model: CCSM3. – J. Climate 19: 2122–
584 2143.
585 Deffontaine, V. et al. 2005. Beyond the Mediterranean peninsulas: evidence of central European
586 glacial refugia for a temperate forest mammal species, the bank vole (*Clethrionomys*
587 *glareolus*). – Mol. Ecol. 14: 1727–1739.
588 Elith, J. and Graham, C.H. 2009. Do they? How do they? WHY do they differ? On finding reasons
589 for differing performances of species distribution models. – Ecography 32: 66–77.
590 Elith, J. et al. 2006. Novel methods improve prediction of species' distributions from occurrence
591 data. – Ecography 29: 129–151.
592 Excoffier, L. et al. 2005. Arlequin ver. 3.0: an integrated software package for population genetics
593 data analysis. – Evol. Bioinform. Online 1: 47–50.
594 Fløjgaard, C. et al. 2009. Ice age distributions of European small mammals: insights from species
595 distribution modelling. – J. Biogeogr. 36: 1152–1163.
596 Fu, Y.X. 1997. Statistical tests of neutrality of mutations against population growth, hitchhiking and
597 background selection. – Genetics 147: 915–925.
598 Guindon, S. and Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large
599 phylogenies by maximum likelihood. – Syst. Biol. 52: 696–704.
600 Guisan, A. and Zimmermann, N.E. 2000. Predictive habitat distribution models in ecology. – Ecol.
601 Model. 135: 147–186.
602 Hall, T.A. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program
603 for Windows 95/98/NT/XP. – Nucleic Acids Symp. Ser. (Oxf.) 41: 95–98.
604 Hasumi, H. and Emori. S. 2004. K-1 coupled GCM (MIROC) description. Center for Climate
605 System Research, University of Tokyo.
606 Heckel, G. et al. 2005. Genetic structure and colonization processes in European populations of
607 the common vole, *Microtus arvalis*. – Evolution 59: 2231–2242.
608 Hewitt, G.M. 2000. The genetic legacy of the Quaternary ice ages. – Nature 405: 907–913.

609 Hijmans, R.J. et al. 2005. Very high resolution interpolated climate surfaces for global land areas. –
610 Int. J. Climatol. 25: 1965–1978.
611 Ho, S.Y.W. et al. 2005. Time dependency on molecular rate estimates and systematic
612 overestimation of recent divergence times. – Mol. Biol. Evol. 22: 1561–1568.
613 Huelsenbeck, J. P. and Ronquist, F. 2001. MrBayes: Bayesian inference of phylogeny. –
614 Bioinformatics 17: 754–755.
615 Hutterer, R. et al. 2008. *Sorex minutus*. – In: IUCN 2009. IUCN Red List of Threatened Species.
616 Version 2009.2 <www.iucnredlist.org>. Downloaded on 12 May 2009.
617 Jaarola, M., and Searle, J.B. 2002. Phylogeography of field voles (*Microtus agrestis*) in Eurasia
618 inferred from mitochondrial DNA sequences. – Mol. Ecol. 11: 2613–2621.
619 Kotlík, P. et al. 2006. A northern glacial refugium for bank voles (*Clethrionomys glareolus*). – Proc.
620 Natl. Acad. Sci. USA 103: 14860–14864.
621 Librado, P. and Rozas, J. 2009. DnaSP ver. 5: a software for comprehensive analysis of DNA
622 polymorphism data. – Bioinformatics 25: 1451–1452.
623 Magri, D. et al. 2006. A new scenario for the Quaternary history of European beech populations:
624 palaeobotanical evidence and genetic consequences. – New Phytol. 171: 199–221.
625 Mascheretti, S. et al. 2003. How did pygmy shrews colonize Ireland? Clues from a phylogenetic
626 analysis of mitochondrial cytochrome *b* sequences. – Proc. Roy. Soc. Lond. B Biol. 270:
627 1593–1599.
628 McDevitt, A.D. et al. 2009. Genetic variation in Irish pygmy shrews *Sorex minutus* (Soricomorpha:
629 Soricidae): implications for colonization history. – Biol. J. Linn. Soc. 97: 918–927.
630 McDevitt, A.D. et al. in press. Postglacial re-colonization of continental Europe by the pygmy shrew
631 (*Sorex minutus*) inferred from mitochondrial and Y chromosomal DNA sequences. – In:
632 Habel J.C. and Assman T. (eds), Relict species – phylogeography and conservation
633 biology. Springer.
634 Mitchell-Jones, A.J. et al. 1999. The Atlas of European Mammals. – Poyser Natural History.
635 Nogués-Bravo, D. 2009. Predicting the past distribution of species climatic niches. – Global Ecol.
636 Biogeogr. 18: 521–531.

637 Nylander, J.A.A. 2004. MrModeltest ver. 2. Program distributed by the author. – Evolutionary
638 Biology Centre, Uppsala University.
639 Ohdachi, S.D. et al. 2006. Molecular phylogenetics of soricid shrews (Mammalia) based on
640 mitochondrial cytochrome *b* gene sequences: with special reference to the Soricinae. – J.
641 Zool. 270: 177–191.
642 Pearson, R.G. et al. 2007. Predicting species distributions from small numbers of occurrence
643 records: a test case using cryptic geckos in Madagascar. – J. Biogeogr. 34: 102–117.
644 Peterson, A.T. et al. 2007. Transferability and model evaluation in ecological niche modeling: a
645 comparison of GARP and Maxent. – Ecography 30: 550–560.
646 Peterson, A.T. et al. 2008. Rethinking receiver operating characteristic analysis applications in
647 ecological niche modeling. Ecol. Model. 213: 63–72.
648 Peterson, A.T. and Nakazawa, Y. 2008. Environmental data sets matter in ecological niche
649 modeling: an example with *Solenopsis invicta* and *Solenopsis richterii*. Glob. Ecol.
650 Biogeogr. 17:135–144.
651 Phillips, S.J. 2008. Transferability, sample selection bias and background data in presence-only
652 modelling: a response to Peterson et al. 2007. – Ecography 31: 272–278.
653 Phillips, S.J. and Dudík, M. 2008. Modeling of species distributions with Maxent: new extensions
654 and a comprehensive evaluation. – Ecography 31: 161–175.
655 Phillips, S.J. et al. 2004. A maximum entropy approach to species distribution modeling. – In:
656 Greiner, R. and Schuurmans, D. (eds), Proc. 21st Int. Conf. on Machine Learning. ACM
657 Press. pp. 655–662.
658 Phillips, S.J. et al. 2006. Maximum entropy modeling of species geographic distributions. Ecol.
659 Model. 190: 231–259.
660 Provan, J. and Bennett, K.D. 2008. Phylogeographic insights into cryptic glacial refugia. – Trends
661 Ecol. Evol. 23: 564–571.
662 Ramos-Onsins, S.E. and Rozas, J. 2002. Statistical properties of new neutrality tests against
663 population growth. – Mol. Biol. Evol. 19: 2092–2100.
664 Ratkiewicz, M. et al. 2002. The evolutionary history of the two karyotypic groups of the common
665 shrew, *Sorex araneus*, in Poland. – Heredity 88: 235–242.

666 Rogers, A.R. and Harpending, H. 1992. Population growth makes waves in the distribution of
667 pairwise genetic differences. – Mol. Biol. Evol. 9: 552–569.
668 Schneider, S. and Excoffier, L. 1999. Estimation of demographic parameters from the distribution
669 of pairwise differences when the mutation rates vary among sites: application to human
670 mitochondrial DNA. – Genetics 152:1079–1089.
671 Sommer, R.S. and Nadachowski, A. 2006. Glacial refugia of mammals in Europe: evidence from
672 fossil records. – Mammal Rev. 36: 251–265.
673 Stewart, J.R. and Lister, A.M. 2001. Cryptic northern refugia and the origins of the modern biota. –
674 Trends Ecol. Evol. 16: 608–613.
675 Stockwell , D.R.B. 1999. Genetic algorithms II. – In: Fielding, A.H. (ed.), Machine Learning
676 Methods for Ecological Applications. Kluwer Academic Publishers, pp. 123–144.
677 Stockwell, D.R.B. and Noble, I.R. 1992. Introduction of sets of rules from animal distribution data: a
678 robust and informative method of data analysis. Math. Comput. Simulat. 33: 385–390.
679 Stockwell, D.R.B. and Peterson, A.T. 2002. Effects of sample size on accuracy of species
680 distribution models. – Ecol. Model. 148: 1–13.
681 Svendsen, J.I. et al. 2004. Late Quaternary ice sheet history of Northern Eurasia. – Quaternary
682 Sci. Rev. 23: 1229–1271.
683 Svenning, J.-C. et al. 2008. Glacial refugia of temperate trees in Europe: insights from species
684 distribution modeling. – J. Ecol. 96: 1117–1127.
685 Swofford, D.L. 2000. PAUP*. Phylogenetic analysis using parsimony (*and other methods), ver. 4.
686 Sinauer.
687 Taberlet, P. et al. 1998. Comparative phylogeography and postglacial colonization routes in
688 Europe. – Mol. Ecol. 7: 453–464.
689 Tajima, F. 1989. Statistical method for testing the neutral mutation hypothesis by DNA
690 polymorphism. – Genetics 123: 585–595.
691 Waltari, E. et al. 2007. Locating Pleistocene refugia: comparing phylogeogaphic and ecological
692 niche model predictions. – *PLoS ONE* 7: e563.
693 Willis, K.J. and van Andel, T.H. 2004. Trees or no trees? The environments of Central and Eastern
694 Europe during the Last Glaciation. – Quaternary Sci. Rev. 23: 2369–2387.

695 Willis, K.J. et al. 2000. The full-glacial forests of Central and South-eastern Europe. – Quaternary
696 Res. 53: 203–213.
697 Wisz, M.S. et al. 2008. Effects of sample size on the performance of species distribution models. –
698 Divers. Distrib. 14: 763–773.
699 Yannic., G. et al. 2008. A new perspective on the evolutionary history of western European *Sorex*
700 *araneus* group revealed by paternal and maternal molecular markers. – Mol. Phyl. Evol. 47:
701 237–250.
702

Figure 1. Bayesian inference tree showing the phylogenetic relationships among *Sorex minutus* samples (*S. volnuchini*, outgroup). Five lineages were found (□ = Pyrenean-Irish, Δ = Italian, ■ = Iberian, ▲= Balkan, and ○ = Northern-Central European). The Northern-Central European lineage is geographically widespread but has not been found within the Southern European peninsulas. Values on branches correspond to Bayesian posterior probabilities. Haplotypes are represented with two-letter country codes followed by an identification number (x2, haplotype frequency = 2 etc.): AD = Andorra, AT = Austria, BY = Belarus, CH = Switzerland, CZ = Czech Republic, DE = Germany, DK = Denmark, ES = Spain, FI = Finland, FR = France, GB = Great Britain, IE = Ireland, IT = Italy, LT = Lithuania, MK = Macedonia, NL = the Netherlands, PL = Poland, RU = Russia, SE = Sweden, SK = Slovakia, TR = Turkey, UA = Ukraine.

Figure 2. Parsimony median joining haplotype network for the Northern-Central European lineage of *Sorex minutus*. Observed haplotypes are shown as grey circles (proportional to frequency) and hypothetical haplotypes are shown as black circles. There is a star-like phylogeny with three central (ancestral) haplotypes. A and B are two central haplotypes from the Netherlands, and C is from Central Ukraine. The dotted black line encircles haplotypes directly linked to A, black lines encircle haplotypes directly linked to B and the dashed line encircles haplotypes directly linked to C (the country of origin for haplotypes is shown next to clusters; two-letters country codes as in Fig. 1). For simplicity, haplotypes from the more diverged Southern European lineages are not shown, but relate to Central-European haplotypes by the addition of several hypothetical haplotypes and > 10 mutational steps. The scale bar represents one mutational step.

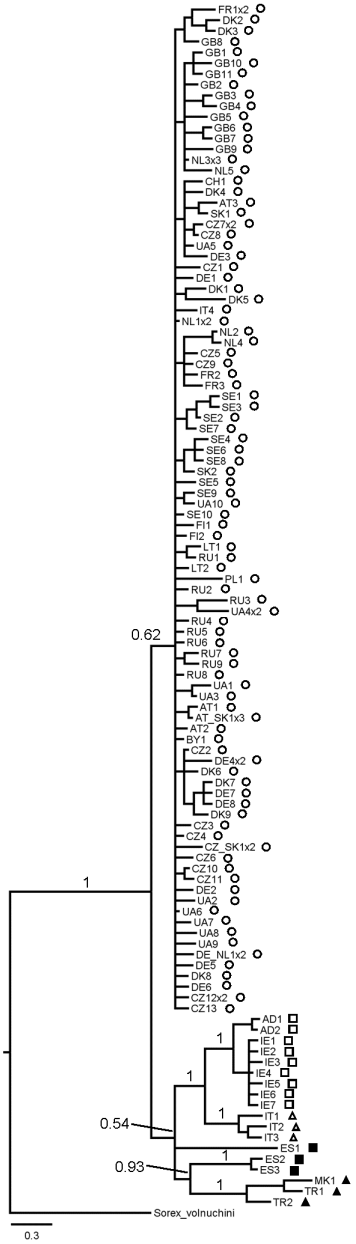
Figure 3. Mismatch distribution for observed (continuous line) and expected (dashed line) pairwise comparisons under a sudden population expansion model among *Sorex minutus* cyt *b* sequences. (a) Mismatch distribution among Eurasian sequences with a bimodal observed distribution where the first peak corresponds to pairwise comparisons among closely related individuals within lineages, while the second peak corresponds to pairwise comparisons among distantly related individuals from different lineages. (b) Mismatch distribution among sequences from the Northern-

731 Central European lineage showing a unimodal distribution, a genetic signature which corresponds
732 to the expected distribution for sudden population expansion.
733
734 Figure 4. Species distribution modelling of *Sorex minutus* in the present and at the Last Glacial
735 Maximum (LGM) using different approaches. Two independent data sets, two algorithms,
736 Maximum entropy (Maxent) and Genetic Algorithm for Rule-set prediction (GARP), and climate
737 reconstructions for the LGM based on two general circulation models (CCSM3 and MIROC3.2)
738 were used. Climatic variables were obtained from WorldClim and two were selected as best
739 predictors with a Jackknife procedure: Annual Mean Temperature and Precipitation of the Warmest
740 Quarter. (a – d) Maxent and GARP modelled present distributions with data sets 1 and 2. (e – l)
741 Maxent and GARP modelled LGM distributions with data sets 1 and 2 using CCSM3 and
742 MIROC3.2. The thick lines (a – d) represent the outline of present-day distribution range of the
743 species, the dark shading corresponds to present-day and LGM suitable climatic conditions, and
744 the white polygon represents the ice extent at the LGM, about 21 Kya (redrawn from Svendsen et
745 al. 2004). Location of samples used for the phylogeographic analysis is shown (lineages as in Fig.
746 1: □ = Pyrenean-Irish, Δ = Italian, ■ = Iberian, ▲= Balkan, and ○ = Northern-Central European).
747

748 Rodrigo Vega, Camilla Fløjgaard, Andrés Lira-Noriega, Yoshinori Nakazawa, Jens-Christian

749 Svenning and Jeremy B. Searle.

750 Figure 1.



751

752

753

754

755

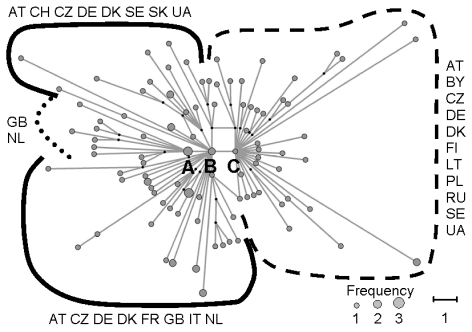
756

757

758 Rodrigo Vega, Camilla Fløjgaard, Andrés Lira-Noriega, Yoshinori Nakazawa, Jens-Christian

759 Svenning and Jeremy B. Searle.

760 Figure 2.



761

762

763

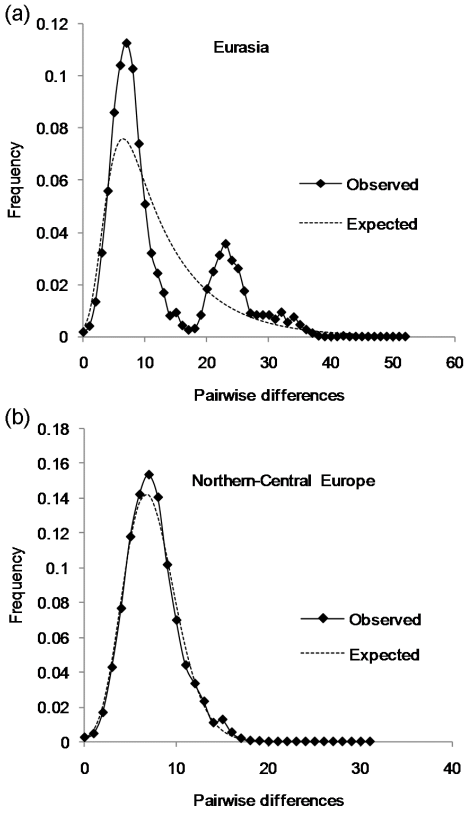
764

765 Rodrigo Vega, Camilla Fløjgaard, Andrés Lira-Noriega, Yoshinori Nakazawa, Jens-Christian

766 Svenning and Jeremy B. Searle.

767 Figure 3.

768



769

770

771

772

Rodrigo Vega, Camilla Fløjgaard, Andrés Lira-Noriega, Yoshinori Nakazawa, Jens-Christian Svenning and Jeremy B. Searle.

Figure 4.